

# A Robust Visual System for Small Target Motion Detection Against Cluttered Moving Backgrounds

Hongxin Wang, Jigen Peng, Xuqiang Zheng and Shigang Yue, *Senior Member, IEEE*

**Abstract**—Monitoring small objects against cluttered moving backgrounds is a huge challenge to future robotic vision systems. As a source of inspiration, insects are quite apt at searching for mates and tracking prey – which always appear as small dim speckles in the visual field. The exquisite sensitivity of insects for small target motion, as revealed recently, is coming from a class of specific neurons called small target motion detectors (STMDs). Although a few STMD-based models have been proposed, these existing models only use motion information for small target detection and cannot discriminate small targets from small-target-like background features (named as fake features). To address this problem, this paper proposes a novel visual system model (STMD+) for small target motion detection, which is composed of four subsystems – ommatidia, motion pathway, contrast pathway and mushroom body. Compared to existing STMD-based models, the additional contrast pathway extracts directional contrast from luminance signals to eliminate false positive background motion. The directional contrast and the extracted motion information by the motion pathway are integrated in the mushroom body for small target discrimination. Extensive experiments showed the significant and consistent improvements of the proposed visual system model over existing STMD-based models against fake features.

**Index Terms**—Visual system model, neural modeling, small target motion detector (STMD), cluttered natural environment, background motion.

## I. INTRODUCTION

**T**HE dynamic visual world is often complex, with many motion cues at different speeds, directions, distances and orientations, exhibiting various physical characteristics such as size, colour, texture and shape [1]–[6]. Being able to detect target motion in the distance and early would put an entity (a robot or an animal) in a good position to prepare for interaction/competition, for example, a flying insect searching for mates in the distance. In the visual world, detecting visual motion in the distance and early often means dealing with small targets with only one or a few pixels in size let alone other physical characteristics. Small target motion detection

has a wide variety of applications in defences, surveillance, security and road safety. However, detecting small targets against cluttered moving backgrounds is always a challenge for artificial visual systems due to limited physical cues of small targets, free motion of camera, and extremely cluttered backgrounds.

How to detect small target motion in cluttered moving backgrounds robustly with limited resources? Research in insects' visual system have revealed one effective solution. Insects show exquisite sensitivity for small target motion [7] and are able to pursue small flying targets with high capture rates [8]. Biological research demonstrates that a class of specific neurons, called small target motion detectors (STMDs), can account for insects' exquisite sensitivity for small target motion [7], [9], [10]. These STMD neurons give peak responses to small targets subtending  $1-3^\circ$  of the visual field, with no response to large bars (typically  $> 10^\circ$ ) or to background movements represented by wide-field grating stimuli [11]. Build a quantitative STMD model is the first step for not only further understanding of the biological visual system but also providing robust and economic solutions of small target detection for an artificial vision system.

The electrophysiological knowledge about the STMD neurons revealed in the past few decades, makes it possible to propose quantitative models, such as elementary small target motion detector (ESTMD) [12] and directionally selective small target motion detector (DSTMD) [13]. Using motion information<sup>1</sup> extracted by large monopolar cells (LMCs) [14], [15], these models are able to detect small moving targets in cluttered backgrounds. However, they cannot discriminate small moving targets from small-target-like background features (as shown in Fig. 1), which means that their detection results may contain a large number of false positives. This is because (1) small-target-like background features are embedded in the cluttered background such as bushes, trees and/or rocks, (2) they are moving with the whole background due to a free flying animal/camera. In this case, these small-target-like features (named as fake features) cannot be simply filtered out by existing STMD-based models with **motion information only** for small target motion detection. To address this problem, other visual information, such as directional contrast<sup>2</sup>, should be combined with motion information for

This work was supported in part by EU HORIZON 2020 Project STEP2DYNA under Grant 691154, in part by EU HORIZON 2020 Project ULTRACEPT under Grant 778062, and in part by the National Natural Science Foundation of China under Grant 11771347. (Hongxin Wang and Jigen Peng contributed equally to this work.) (Corresponding author: Shigang Yue.)

H. Wang and S. Yue are with the Machine Life and Intelligence Research Center, Guangzhou University, Guangzhou 510006, China, and also with the Computational Intelligence Lab, School of Computer Science, University of Lincoln, Lincoln LN6 7TS, U.K. (email: syue@lincoln.ac.uk).

J. Peng is with the School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China (email: jg-peng@gzhu.edu.cn).

X. Zheng is with the Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100029, China.

<sup>1</sup>Motion information refers to luminance changes of a pixel with respect to time. From the view of mathematics, it is equivalent to temporal derivative of a pixel.

<sup>2</sup>Directional contrast denotes luminance changes of a pixel along different spatial directions. From the view of mathematics, it corresponds to directional derivatives of a pixel.



Fig. 1. A small target is moving in the cluttered natural background which contains a number of small-target-like features (or called fake features). The small target and fake features all appear as small dim speckles whose sizes vary from one pixel to a few pixels, since they are far away from the animal/camera.

distinguishing small targets from fake features.

In the insects' visual systems, multiple visual cues are extracted by different specialized neural circuits [16]–[18]. Multiple neural circuits could be coordinated to discriminate small target motion. For example, in the lamina layer, large monopolar cells (LMCs) [14], [15] have been described as temporal band-pass filters which extract motion information from luminance signals [12], [13], [19]; and amacrine cells (AMCs) [20]–[22] linked to multi adjacent ommatidia with thin extending fibers, may constitute a contrast pathway with their downstream neurons to extract directional contrast from luminance signals. Although the contribution from the AMCs to STMD neural circuits in insects is unknown, it is clear that with directional contrast and motion information together, an artificial vision system could discriminate small moving targets from fake features robustly.

Inspired by the above biological findings, this paper proposes a new visual system model (STMD+) to detect small target motion in cluttered moving backgrounds. The main contribution of this work is combining motion information with directional contrast to successfully discriminate small targets from fake features. The rest of this paper is organized as follows. Section II reviews related work on small target motion detection. In Section III, we introduce our proposed visual system model. Section IV provides extensive performance evaluation as well as comparisons against the existing models. Finally, we conclude this paper in Section V.

## II. RELATED WORK

Small target motion detection aims to detect objects of interest which move against cluttered natural environments and appear as small dim speckles<sup>3</sup> in images. Inspired by the insect's motion-sensitive neurons, several models have been developed to detect small target motion. In this section, we firstly review motion-sensitive neural models, then briefly discuss traditional motion detection and small target detection approaches.

### A. Motion-sensitive Neural Models

Small target motion detectors (STMDs) [9]–[11] and lobula plate tangential cells (LPTCs) [23], [24] are widely inves-

<sup>3</sup>The sizes of small dim speckles vary from 1 pixel to  $10 \times 10$  pixels, whereas other physical characteristics, such as color, shape and texture, are difficult to recognize and cannot be used for motion detection.

tigated motion-sensitive neurons, where the former shows exquisite sensitivity to small target motion while the latter responds strongly to wide-field motion.

Wiederman *et al.* [12] presented a mathematical model called ESTMD to simulate STMD neurons. It can detect the presence of small moving targets, but is unable to estimate motion direction. To address this issue, directional selectivity has been introduced into the ESTMD [13], [19], [25]. However, these models cannot discriminate small targets from fake features, as they only make use of motion information.

The first LPTC model called elementary motion detector (EMD) [26], is originally inferred from the insects' behavior. Following that, several studies have been done to further improve the EMD, such as [27]–[29]. These models can detect all objects' motion, nevertheless they are unable to distinguish small moving objects from large ones.

### B. Traditional Motion Detection Methods

Traditional motion detection methods such as optical flow [30], background subtraction [31] and temporal differencing [32], have been developed to detect normal-sized objects like pedestrians and vehicles. They utilize physical characteristics including shape, color and texture, to segment regions corresponding to moving objects from the background. Nonetheless, these methods would be powerless for objects that are as small as one pixel or a few pixels, because it is difficult to identify objects' physical characteristics in such small sizes. Additionally, the above-mentioned methods may not work for cluttered moving backgrounds, as small moving objects could be submerged among the pixel error when applying background motion compensation [33].

### C. Infrared Small Target Detection

Previous research and application of small target detection has mainly focused on infrared images [34]–[36]. These infrared-based methods strongly rely on significant temperature differences between the background and objects of interest, such as rockets, jets and missiles. However, such significant temperature difference is rare in natural world. Moreover, the detection environment of these methods were mainly sky and/or ocean, which are much more clear and homogeneous than the cluttered natural environments. These infrared-based methods may not work in a natural environment with lots of bushes, trees, sunlight and shadows, let alone to meet the needs of compact in size and low energy consumption in real applications [37]–[41].

## III. FORMULATION OF THE SYSTEM

In this section, we first illustrate the proposed visual system model schematically, then elaborate on its components in following subsections.

The proposed visual system model is composed of four subsystems, including ommatidia, motion pathway, contrast pathway and mushroom body [42], [43], as illustrated in Fig. 2(a). The luminance signals are received and smoothed by the ommatidia, then applied to the motion and contrast pathways.

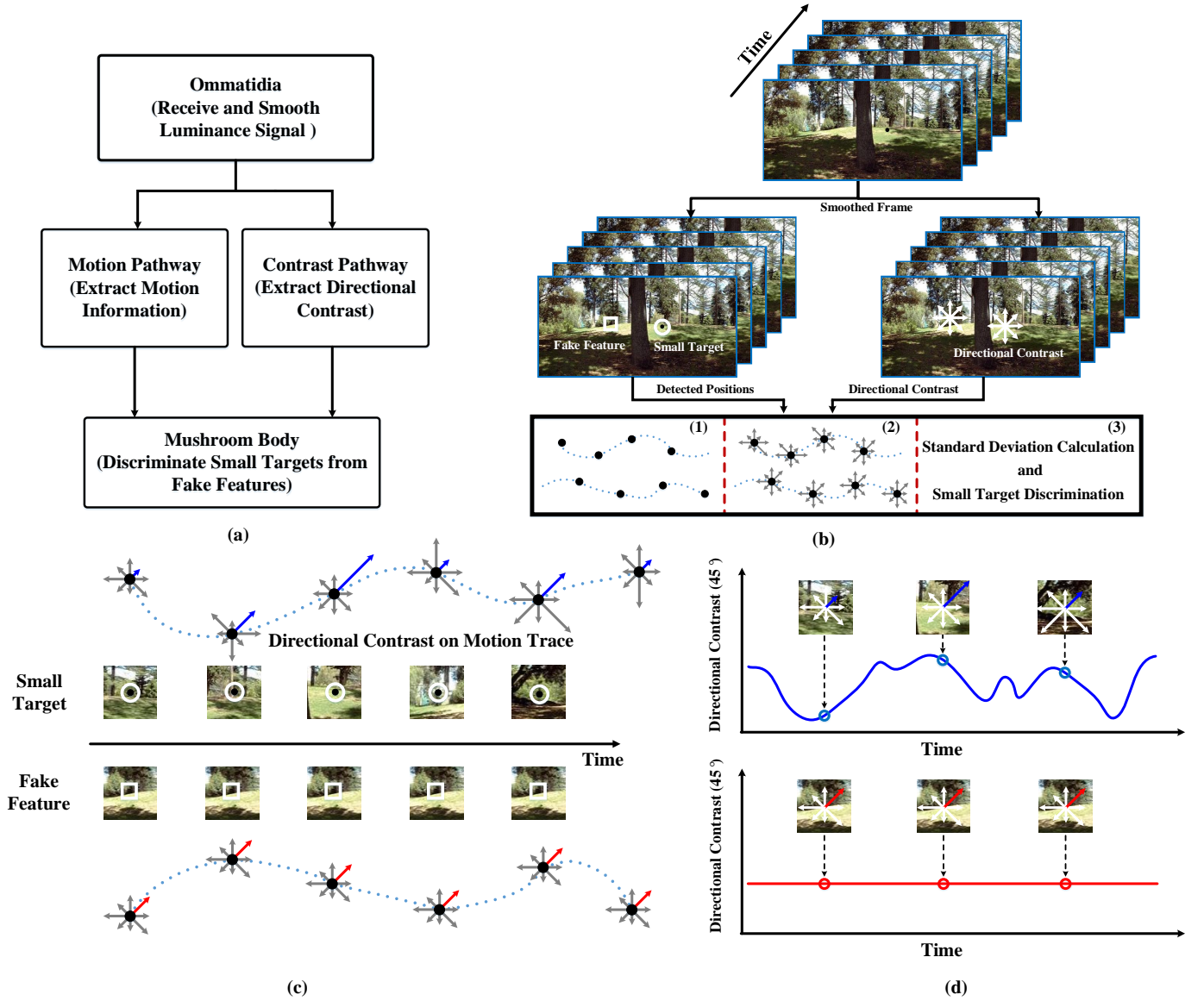


Fig. 2. (a) Schematic illustration of the proposed visual system model (STMD+). (b) Image processing of the proposed visual system model. (c) Directional contrast on two motion traces which are caused by the small target and fake feature, respectively. Directional contrast is denoted by arrows along different directions where the arrow's length represents the strength of the directional contrast. For the small target (top), its directional contrast varies significantly with time. However, for the fake feature (bottom), its directional contrast shows little change over time. (d) Directional contrast along  $45^\circ$  direction of the small target (top) and fake feature (bottom) with respect to time.

These two pathways separately extract motion information and directional contrast which are finally integrated in the mushroom body to discriminate small targets from fake features.

Fig. 2(b) shows the image processing of the proposed visual system model, where the input image sequence is processed frame by frame. In each frame, both small targets and fake features are located by computing luminance changes of each pixel over time, while directional contrast is obtained by calculating luminance changes of each pixel along different directions. The detected positions and directional contrast are further processed as follows.

- 1) Successively record the detected positions to infer motion traces.
- 2) Extract the directional contrast on each motion trace.
- 3) Compute the standard deviation of directional contrast

on each motion trace and compare it with a threshold for distinguishing small targets from fake features.

Our motivation is mainly based on the following observations: the directional contrast of small targets varies significantly with time, since they have relative movement to the background; on the contrary, the directional contrast of fake features shows little change over time, as they are static relative to the background. The variation amount in the directional contrast with time is represented by the standard deviation, which is taken as the criterion for small target discrimination. Fig. 2(c) visually displays the directional contrast on two typical motion traces that are separately caused by the small target and fake feature. As an example, Fig. 2(d) presents the directional contrast along  $45^\circ$  direction, which is used to calculate the standard deviation for this direction.

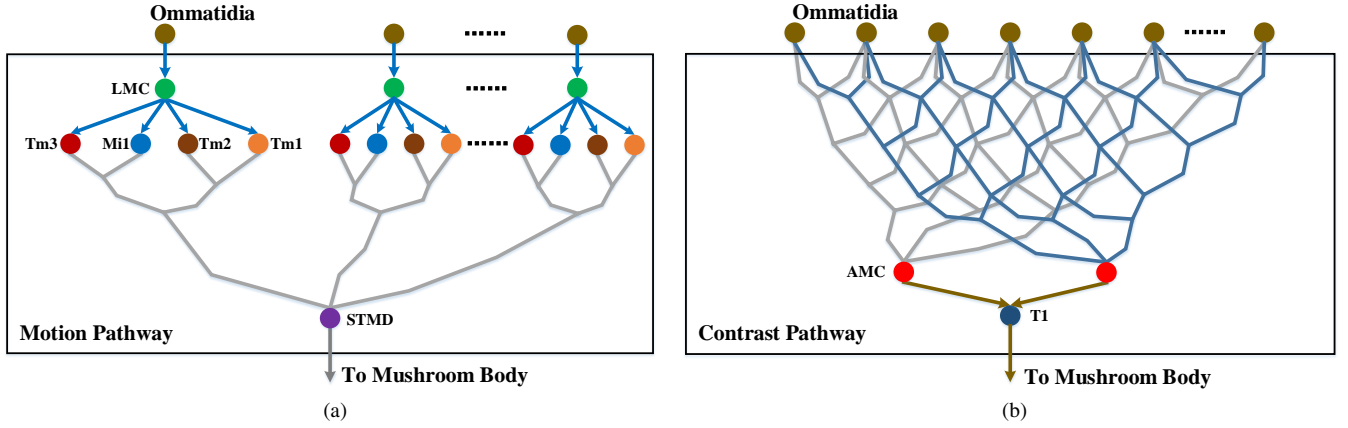


Fig. 3. Wiring sketches of motion and contrast pathways. In subplots, each colored node denotes a neuron. For clear illustration, only one STMD and T1 neurons are presented here. (a) Motion pathway. (b) Contrast pathway. Note that each AMC collects signals from multiple ommatidia while each LMC receives signals from a single ommatidium.

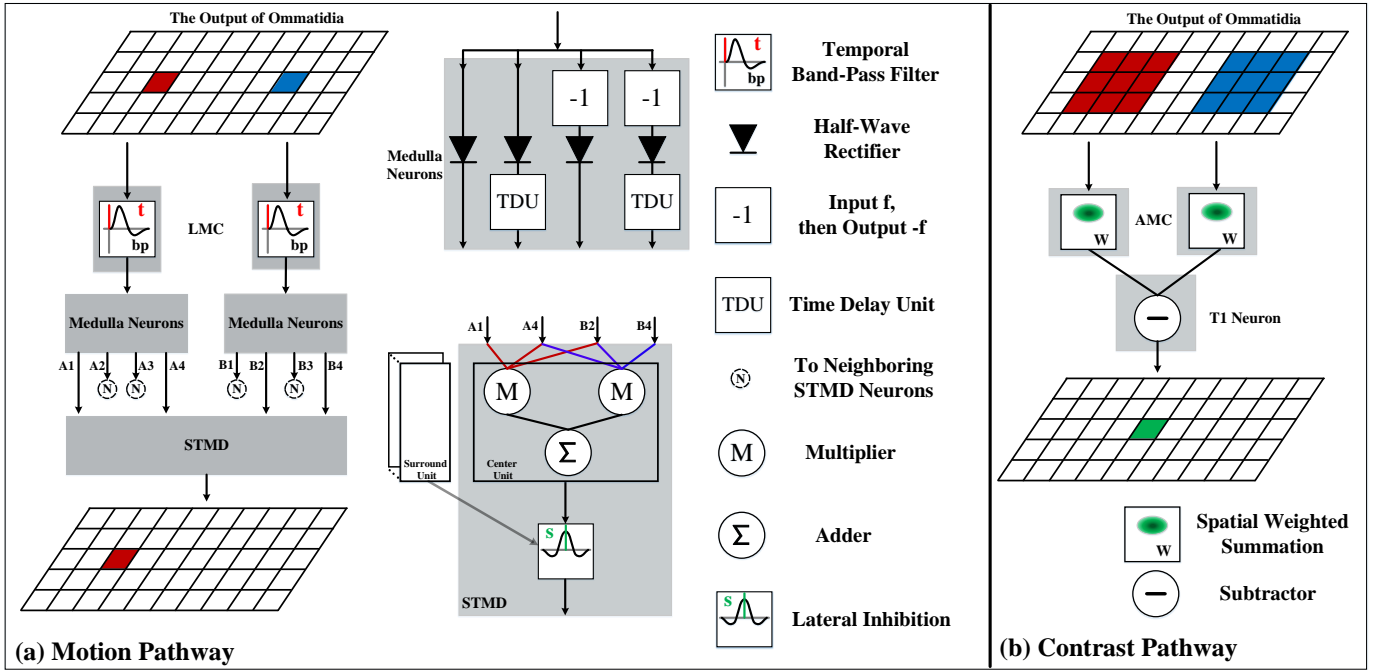


Fig. 4. Schematic illustration of models of motion and contrast pathways. For clear illustration, only one STMD and T1 neurons are presented here. However, these types of neurons are all arranged in matrix form in the proposed visual system.

#### A. Ommatidia

Ommatidia act as luminance receptors to perceptive visual stimuli from the natural world [44]. In the proposed visual system, they are arranged in a matrix and modelled as spatial Gaussian filters, each of which captures and smooths the luminance of each pixel in the input image. Formally, let  $I(x, y, t) \in \mathbb{R}$  denote the input image sequence, where  $x, y$  and  $t$  are spatial and temporal field positions. The output of an ommatidium  $P(x, y, t)$  is given by,

$$P(x, y, t) = \iint I(u, v, t) G_{\sigma_1}(x - u, y - v) du dv \quad (1)$$

where  $G_{\sigma_1}(x, y)$  is a Gaussian function, defined as

$$G_{\sigma_1}(x, y) = \frac{1}{2\pi\sigma_1^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_1^2}\right). \quad (2)$$

#### B. Motion Pathway

As shown in Fig. 3(a), the motion pathway consists of large monopolar cells (LMCs) [14], [15], medulla neurons (i.e., Mi1, Tm1, Tm2 and Tm3) [45], [46], small target motion detectors (STMDs) [9]–[11]. The output of ommatidia is first fed into LMCs, then processed by medulla neurons and finally integrated by STMDs. Fig. 4(a) displays the model of the motion pathway, which is elaborated as follows.

**1) Large Monopolar Cells (LMCs):** Objects' motion can induce luminance changes of pixels with time. These luminance changes are extracted by the LMCs, each of which is modelled by a temporal band-pass filter that is defined as the



difference of two Gamma kernels (see Fig. 4(a)). That is,

$$H(t) = \Gamma_{n_1, \tau_1}(t) - \Gamma_{n_2, \tau_2}(t) \quad (3)$$

$$\Gamma_{n, \tau}(t) = (nt)^n \frac{\exp(-nt/\tau)}{(n-1)! \cdot \tau^{n+1}} \quad (4)$$

where  $H(t)$  denotes the impulse response of the band-pass filter,  $\Gamma_{n, \tau}(t)$  stands for the Gamma kernel [47],  $n$  and  $\tau$  refers to the order and time constant of the Gamma kernel  $\Gamma_{n, \tau}(t)$ . Then the output of each LMC can be calculated by convolving  $H(t)$  with the output of ommatidia  $P(x, y, t)$ ,

$$L(x, y, t) = \int P(x, y, s)H(t-s)ds. \quad (5)$$

The  $L(x, y, t)$  reflects luminance changes of pixel  $(x, y)$  over time  $t$ , where a positive  $L(x, y, t)$  means luminance increase while a negative  $L(x, y, t)$  suggests luminance decrease.

**2) Medulla Neurons:** Medulla neurons including Tm1, Tm2, Tm3 and Mi1, constitute four parallel channels to process the output of LMCs  $L(x, y, t)$ . The Tm3 and Tm2 are modelled as half-wave rectifiers to separate  $L(x, y, t)$  into luminance increase and decrease components. Let  $S^{\text{Tm3}}(x, y, t)$  and  $S^{\text{Tm2}}(x, y, t)$  denote the output of the Tm3 and Tm2, respectively, then they are given by

$$S^{\text{Tm3}}(x, y, t) = [L(x, y, t)]^+ \quad (6)$$

$$S^{\text{Tm2}}(x, y, t) = [-L(x, y, t)]^+ \quad (7)$$

where  $[x]^+$  denotes  $\max(x, 0)$ . The Mi1 and Tm1 further temporally delay  $S^{\text{Tm3}}(x, y, t)$  and  $S^{\text{Tm2}}(x, y, t)$  by convolving them with a Gamma kernel. That is,

$$S^{\text{Mi1}}_{(n, \tau)}(x, y, t) = \int [L(x, y, s)]^+ \cdot \Gamma_{n, \tau}(t-s)ds \quad (8)$$

$$S^{\text{Tm1}}_{(n, \tau)}(x, y, t) = \int [-L(x, y, s)]^+ \cdot \Gamma_{n, \tau}(t-s)ds \quad (9)$$

where  $S^{\text{Mi1}}_{(n, \tau)}(x, y, t)$  and  $S^{\text{Tm1}}_{(n, \tau)}(x, y, t)$  represent the outputs of the Mi1 and Tm1, respectively;  $n$  and  $\tau$  are the order and time constant of the Gamma kernel, which separately determine the order and time-delay length of the time delay unit (TDU) (see Fig. 4(a)).

**3) Small Target Motion Detectors (STMDs):** As can be seen from Fig. 4(a), each STMD collects the outputs of medulla neurons located at two pixels, i.e.,  $(x, y)$  and  $(x'(\theta), y'(\theta))$  which are defined as

$$\begin{aligned} x'(\theta) &= x + \alpha_1 \cos \theta \\ y'(\theta) &= y + \alpha_1 \sin \theta \end{aligned} \quad (10)$$

where  $\alpha_1$  is a constant,  $\theta$  denotes the preferred direction of the STMD. When a dim object successively moves over pixels  $(x, y)$  and  $(x'(\theta), y'(\theta))$ , a luminance decrease followed by a luminance increase will appear at each of these two pixels. These luminance increase and decrease signals are first aligned in time domain and then multiplied together so as to produce a large response [13]. That is,

$$\begin{aligned} D(x, y, t, \theta) &= S^{\text{Tm3}}(x, y, t) \cdot \left\{ S^{\text{Tm1}}_{(n_4, \tau_4)}(x, y, t) \right. \\ &\quad \left. + S^{\text{Mi1}}_{(n_3, \tau_3)}(x'(\theta), y'(\theta), t) \right\} \cdot S^{\text{Tm1}}_{(n_5, \tau_5)}(x'(\theta), y'(\theta), t) \end{aligned} \quad (11)$$

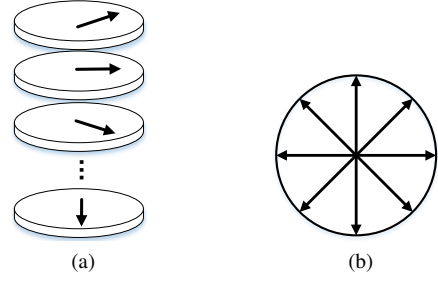


Fig. 5. (a) Illustration of neurons which are located at the same position, but have different preferred directions. The black arrows denote preferred directions. (b) Illustration of different preferred directions in the x-y plane.

where  $D(x, y, t, \theta)$  denotes the output of the STMD neuron with a preferred direction  $\theta$ . Here,  $\theta$  belongs to  $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$ , corresponding to eight preferred directions of STMD neurons (see Fig. 5). It is worthy to note that  $\tau_3, \tau_4$  and  $\tau_5$  are determined by the different delays among the luminance changes, while  $n_3, n_4$  and  $n_5$  are accordingly tuned to guarantee appropriate Gamma kernel shapes [13].

So far, the obtained  $D(x, y, t, \theta)$  can detect both small and large moving objects in the forms of producing a large response. In order to suppress the responses to large moving objects, the  $D(x, y, t, \theta)$  is further laterally inhibited by convolving with an inhibition kernel  $W_s(x, y)$ . That is,

$$E(x, y, t, \theta) = \iint D(u, v, t, \theta) W_s(x-u, y-v) du dv \quad (12)$$

where  $E(x, y, t, \theta)$  represents the inhibited signal; the inhibition kernel  $W_s(x, y)$  is defined as

$$W_s(x, y) = A \cdot [g(x, y)]^+ + B \cdot [g(x, y)]^- \quad (13)$$

$$g(x, y) = G_{\sigma_2}(x, y) - e \cdot G_{\sigma_3}(x, y) - \rho \quad (14)$$

where  $[x]^+$  and  $[x]^-$  respectively denote  $\max(x, 0)$  and  $\min(x, 0)$ ;  $A, B, e$  and  $\rho$  are constant.

By comparing the  $E(x, y, t, \theta)$  with a detection threshold  $\beta$ , we can find the positions of small moving objects. Specially, if  $E(x, y, t, \theta) > \beta$ , then we believe that a small object moving along direction  $\theta$  is located at pixel  $(x, y)$  and time  $t$ . However, it cannot distinguish small targets and fake features that can be both recognized as small moving objects. To address this issue, we construct a contrast pathway accounting for directional contrast calculation.

### C. Contrast Pathway

As shown in Fig. 3(b), contrast pathway is composed of amacrine cells (AMCs) [20]–[22] and T1 neurons [48], [49]. The output of ommatidia is firstly fed into AMCs, then processed by T1 neurons. Fig. 4(b) displays the model of the contrast pathway, which is elaborated as follows.

**1) Amacrine Cells (AMCs):** Each AMC receives the output of multiple ommatidia located in a small region and serves as a weighted summation unit, as presented in Fig. 4(b). Here, we define the weight function as

$$W_A(x, y) = \frac{1}{2\pi\eta^2} \exp\left(-\frac{x^2 + y^2}{2\eta^2}\right) \quad (15)$$

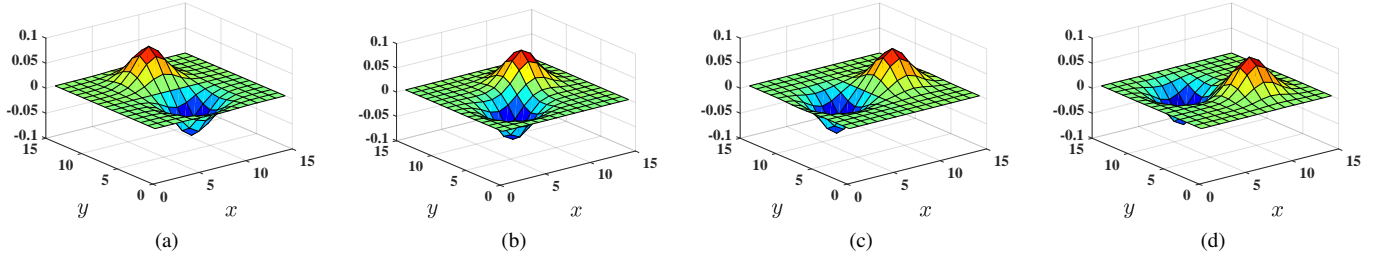


Fig. 6. Illustration of the convolution kernel  $W_T(x, y, \phi)$ . (a)  $\phi = 0$ . (b)  $\phi = \frac{\pi}{4}$ . (c)  $\phi = \frac{\pi}{2}$ . (d)  $\phi = \frac{3\pi}{4}$ .

where  $\eta$  is constant. Then the output of each AMC  $A(x, y, t)$  can be given by

$$A(x, y, t) = \iint P(u, v, t) W_A(x - u, y - v) du dv \quad (16)$$

where  $P(x, y, t)$  is the output of ommatidia defined in (1).

**2) T1 Neurons:** The T1 neuron layer is adopted to extract the directional contrast along different directions. The directional contrast at  $(x, y)$  along direction  $\phi$  is defined as the difference between the outputs of two AMCs that are located at  $(x + \alpha_2 \cos \phi, y + \alpha_2 \sin \phi)$  and  $(x - \alpha_2 \cos \phi, y - \alpha_2 \sin \phi)$ . Here,  $\alpha_2$  is a constant. Let  $T(x, y, t, \phi)$  denote the output of a T1 neuron with a preferred direction  $\phi$ , then it can be given by

$$T(x, y, t, \phi) = A(x + \alpha_2 \cos \phi, y + \alpha_2 \sin \phi, t) - A(x - \alpha_2 \cos \phi, y - \alpha_2 \sin \phi, t). \quad (17)$$

Substituting (16) in (17), we have

$$T(x, y, t, \phi) = \iint P(u, v, t) W_T(x - u, y - v, \phi) du dv \quad (18)$$

where the convolution kernel  $W_T(x, y, \phi)$  represents

$$W_T(x, y, \phi) = W_A(x + \alpha_2 \cos \phi, y + \alpha_2 \sin \phi) - W_A(x - \alpha_2 \cos \phi, y - \alpha_2 \sin \phi). \quad (19)$$

Here  $\phi$  belongs to  $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ , corresponding to four preferred directions of T1 neurons. It is worthy to note that the convolution kernel  $W_T(x, y, \phi)$  is one of the directional derivative operators [50], [51], which can extract anisotropic luminance variations (see Fig. 6).

### D. Mushroom Body

In the proposed visual system, the mushroom body [42], [43] receives two types of neural outputs, including the output of STMDs  $E(x, y, t, \theta)$  and the output of T1 neurons  $T(x, y, t, \phi)$ . These neural outputs are integrated to discriminate small targets from fake features via the following three procedures.

**1) Motion Trace Recording:** The output of STMDs  $E(x, y, t, \theta)$  is employed to record motion traces of small objects. For a detection threshold  $\beta$  and a starting time  $t_0$ , if there exists a pixel A  $(x_A, y_A)$  and motion direction  $\theta_A$  which satisfy  $E(x_A, y_A, t_0, \theta_A) > \beta$ , then we believe that a small object<sup>4</sup> is detected at pixel  $(x_A, y_A)$  and its motion direction

<sup>4</sup>The detected object could be a small target or a fake feature, which cannot be discriminated by the STMDs.

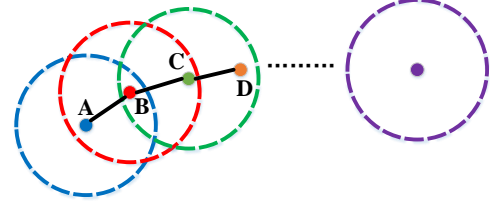


Fig. 7. Motion trace recording. Each node denotes a detected pixel while each circle represents a small neighborhood. If pixel B is the nearest detected point to pixel A, and pixel B is in the neighborhood of pixel A, then we believe that pixels A and B belong to the same motion trace. Repeating this step, a motion trace could be recorded.

is  $\theta_A$ . Similarly, at next time step  $t_1$ , another pixel B  $(x_B, y_B)$  and motion direction  $\theta_B$  can be detected. Especially, if pixel B  $(x_B, y_B)$  is the nearest detected point to pixel A  $(x_A, y_A)$ , and pixel B is in the small neighborhood of pixel A, then we believe that pixels A and B belong to the same motion trace denoted by  $TR$ . Repeating the above steps, the motion trace  $TR$  can be recorded during a time period, as shown in Fig. 7. The  $TR$  can be described as,

$$TR = (x(t), y(t), \theta(t)), t \in [t_0, t_n] \quad (20)$$

where  $x(t)$  and  $y(t)$  represent x and y coordinates at time  $t$ ,  $\theta(t)$  denotes motion direction,  $t_0$  and  $t_n$  are the starting time and current time.

**2) Information Integration:** Once motion traces are recorded, we can obtain their directional contrast by substituting (20) into  $T(x, y, t, \phi)$ . That is,

$$Q(t, \phi) = T(x(t), y(t), t, \phi), t \in [t_0, t_n] \quad (21)$$

where  $Q(t, \phi)$  denotes the directional contrast along direction  $\phi$  on the motion trace  $TR$ ;  $(x(t), y(t))$  stands for the point on the motion trace. To quantify the variation amount in the directional contrast, we calculate the standard deviation ( $SD$ ) of the  $Q(t, \phi)$  during a time period  $[t_{n-m}, t_n]$ , denoted by  $SD(t_{n-m}, t_n, \phi)$ . Here  $m$  represents the sample number for the  $SD$  calculation.

**3) Small Target Discrimination:** We determine whether a detected object is a small target or a fake feature, using the standard deviations of the directional contrast on the object's motion trace, i.e.,  $SD(t_{n-m}, t_n, \phi)$ . If the  $SD(t_{n-m}, t_n, \phi)$  is smaller than a certain threshold, we believe that the detected object is a fake feature; Otherwise, it is a small target.

TABLE I  
PARAMETERS OF THE PROPOSED VISUAL SYSTEM MODEL.

Eq.	Parameters
(1)	$\sigma_1 = 1$
(3)	$n_1 = 2, \tau_1 = 3, n_2 = 6, \tau_2 = 9$
(10)	$\alpha_1 = 3$
(11)	$n_3 = 3, \tau_3 = 15, n_4 = 5, \tau_4 = 25, n_5 = 8, \tau_5 = 40$
(13)	$A = 1, B = 3$
(14)	$\sigma_2 = 1.5, \sigma_3 = 3.0, e = 1, \rho = 0$
(15)	$\eta = 1.5$
(17)	$\alpha_2 = 3$

### E. Parameter Setting

Parameters of the proposed visual system model are listed in Table I, where the parameters of the motion pathway are determined by the analysis in [13] while those of the contrast pathway are tuned empirically. These parameters are chosen to satisfy the functionality, which are mainly determined by the velocity and size ranges of the moving targets. They will not be changed in the following experiments unless stated.

The proposed visual system model is written in Matlab (The MathWorks, Inc., Natick, MA). The computer used in the experiments is a standard laptop with a 2.50GHz Intel Core i7 CPU and 16GB DDR3 memory. The source code can be found at <https://github.com/wanghongxin/STMD-Plus>.

## IV. RESULTS AND DISCUSSIONS

The proposed visual system model is evaluated on a synthetic dataset [52] and a real dataset (STNS dataset) [19]. The synthetic dataset contains a number of image sequences which are synthesized by using real background images and a computer generated small target (a black block). These image sequences all display the motion of the small target against the cluttered moving backgrounds, which are different in the target sizes, target velocities, background velocities, background types and so on. The sampling frequencies of the synthetic videos are all equal to 1000 Hz. The STNS dataset is a collection of 25 real videos featuring various moving targets and environments. The scenarios include many kinds of challenges, such as heavy clutter, camera motion and changes in overall brightness. The STNS dataset (videos and manual ground truth annotations) is available at [https://figshare.com/articles/STNS\\_Dataset/4496768](https://figshare.com/articles/STNS_Dataset/4496768).

To quantitatively evaluate the detection performance, two metrics are defined as following [34],

$$D_R = \frac{\text{number of true detections}}{\text{number of actual targets}} \quad (22)$$

$$F_A = \frac{\text{number of false detections}}{\text{number of images}} \quad (23)$$

where  $D_R$  and  $F_A$  denote detection rate and false alarm rate, respectively. The detected result is considered correct if the pixel distance between the ground truth and the result is within a threshold (5 pixels).



Fig. 8. Input frame at time  $t_0 = 1000$  ms whose resolution is 500 pixels (in horizontal) by 250 pixels (in vertical). The small target (the black block) and the cluttered background are moving from left to right. Their velocities are all equal to 250 pixel/s, where arrow  $V_B$  denotes the motion direction of the background. The tree which is regarded as a large object, is also moving due to the background motion.

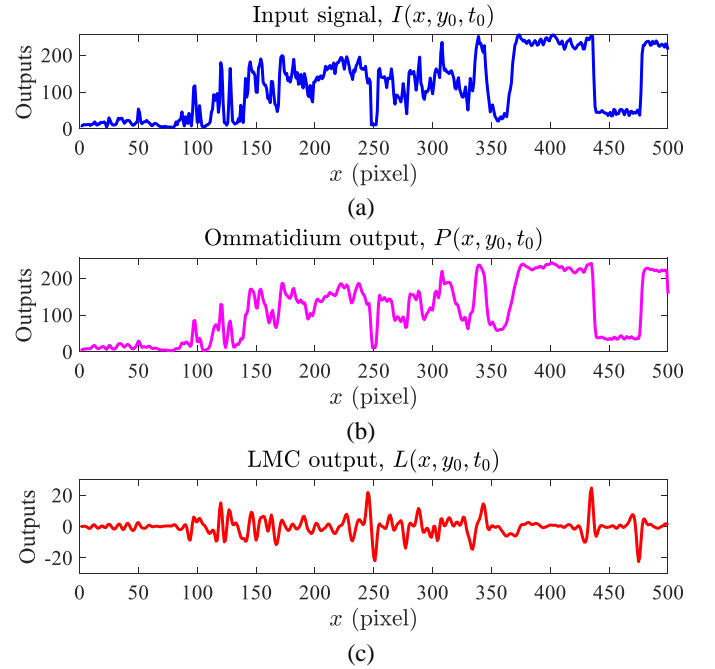


Fig. 9. In each subplot, the horizontal axis denotes  $x$  coordinate while the vertical axis represents neural outputs. (a) Input luminance signal  $I(x, y_0, t_0)$ . (b) Ommatidium output  $P(x, y_0, t_0)$ . (c) LMC output  $L(x, y_0, t_0)$ .

### A. Signal Processing in the Motion Pathway

To intuitively illustrate the signal processing in the motion pathway, we observe the output of each neural layer with respect to  $x$  by setting  $y$  and  $t$  as  $y_0 = 125$  pixel and  $t_0 = 1000$  ms. Fig. 8 shows the input frame at time  $t_0 = 1000$  ms, where the luminance signal  $I(x, y_0, t_0)$  on the middle line is presented in Fig. 9(a). Its resulting ommatidium output and LMC output are displayed in Fig. 9(b) and (c), respectively. The ommatidium output is a smoothed version of the input signal. The LMC output reveals the luminance changes of pixels, where the positive values correspond to luminance increase while the negative values suggest luminance decrease.

Fig. 10(a) demonstrates the four inputs of the STMDs when the preferred direction  $\theta$  is set to 0. Specifically, the  $S^{\text{Tm3}}(x, y_0, t_0)$  is the positive part of the LMC output; the  $S^{\text{M1}}_{(n_3, \tau_3)}(x + \alpha_1, y_0, t_0)$  denotes the delayed version of the positive part of the LMC output with a shift of  $\alpha_1$  pixels; the

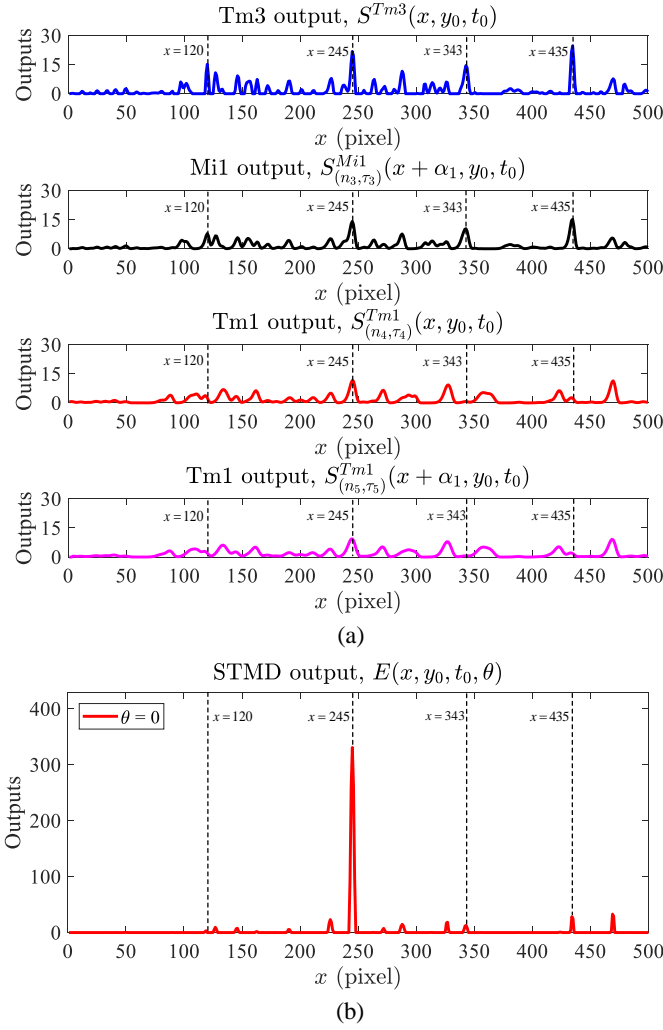


Fig. 10. In each subplot, the horizontal axis denotes  $x$  coordinate while the vertical axis represents neural outputs. (a) Four inputs of the STMDs when the preferred direction  $\theta$  is set to 0, i.e.,  $S^{Tm3}(x, y_0, t_0)$ ,  $S^{Ml1}_{(n3, \tau3)}(x + \alpha_1, y_0, t_0)$ ,  $S^{Tm1}_{(n4, \tau4)}(x, y_0, t_0)$  and  $S^{Tm1}_{(n5, \tau5)}(x + \alpha_1, y_0, t_0)$ . (b) STMD output  $E(x, y_0, t_0, \theta)$  when the preferred direction  $\theta$  is equal to 0.

$S^{Tm1}_{(n4, \tau4)}(x, y_0, t_0)$  stands for the delayed version of the negative part of the LMC output; and the  $S^{Tm1}_{(n5, \tau5)}(x + \alpha_1, y_0, t_0)$  represents the delayed version of negative part of the LMC output with a shift of  $\alpha_1$  pixels. Fig. 10(b) further shows the output of STMDs, where a high response appears at the position of the small target ( $x = 245$ ) while the responses at other positions are effectively suppressed. This is because the four peaks located at the position of the small target are aligned (see Fig. 10(a)), which will produce a strong response after the multiplication, summation and lateral inhibition in the STMD (see Fig. 4). For other positions e.g.,  $x = 120, 343, 435$ , the peaks on the four curves exhibit a low aligning probability, hence producing a weak response. Note that the lateral inhibition is introduced to suppress the responses to large objects, such as the tree displayed in Fig. 8.

It is worthy to note that the above analysis is based on the presetting of the preferred direction  $\theta = 0$ . When we change the preferred direction  $\theta$ , different STMD outputs can be

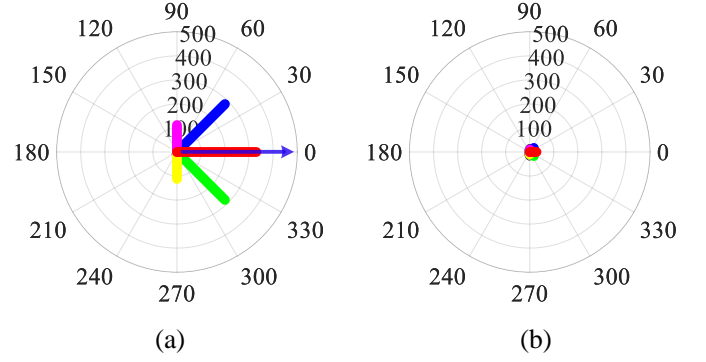


Fig. 11. In the polar coordinate system, the angular coordinate represents the preferred direction  $\theta$  while the radial coordinate denotes the STMD output. (a) STMD outputs at position  $x = 245$  along eight preferred directions  $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$ . The blue arrow stands for the motion direction of the small target. (b) STMD outputs at position  $x = 435$  along eight preferred directions  $\theta$ .

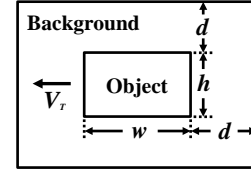


Fig. 12. External rectangle and neighboring background rectangle of an object. Arrow  $V_r$  denotes the motion direction of the object.  $w$  represents object width while  $h$  stands for object height.

calculated. Fig. 11 presents the STMD outputs at the positions  $x = 245$  and  $x = 435$  along eight preferred directions  $\theta$ , where  $x = 245$  is the position of the small target and  $x = 435$  corresponds to the position of the large tree. As shown in Fig. 11(a), for the small target, the STMD shows strong directional selectivity. As the preferred direction deviates from the motion direction of the small target, the STMD output will decrease correspondingly. On the other hand, the direction of the small target can be estimated by computing the summation of these output vectors [13]. For the large tree (see Fig. 11(b)), the outputs of the STMD along eight preferred directions are very low, suggesting that the STMD is not interested in large moving objects.

### B. Characteristics of the STMD

To further demonstrate the characteristics of the STMD, we compare its outputs to objects with different velocities, widths, heights and Weber contrast. As shown in Fig. 12, width (or height) represents object length extended parallel (or orthogonal) to the motion direction. Weber contrast is defined by the following equation,

$$\text{Weber contrast} = \frac{|\mu_t - \mu_b|}{255} \quad (24)$$

where  $\mu_t$  is the average pixel value of the object, while  $\mu_b$  is the average pixel value in neighboring area around the object. If the size of a object is  $w \times h$ , the size of its background rectangle is  $(w + 2d) \times (h + 2d)$ , where  $d$  is a constant which equals to 10 pixels. The initial Weber contrast, velocity, width



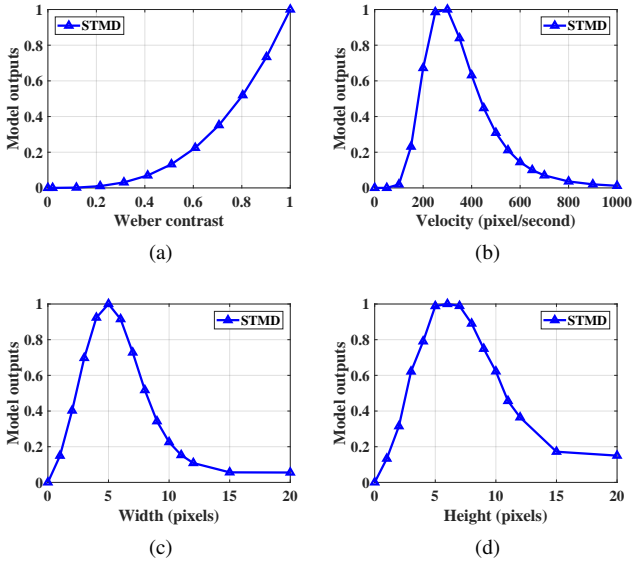


Fig. 13. STMD outputs to moving objects with different Weber contrast, velocities, widths and heights. (a) Different Weber contrast. (b) Different velocities. (c) Different widths. (d) Different heights.

and height of the object are set as 1, 250 pixel/s, 5 pixels and 5 pixels, respectively.

Fig. 13(a) shows the STMD output with respect to the Weber contrast. As can be seen, the STMD output increases as the increase of Weber contrast, until reaches maximum at Weber contrast = 1. This indicates that the higher Weber contrast of an object is, the easier it can be detected. Fig. 13(b) presents the STMD output with regard to the velocity of the moving object. Obviously, the STMD output peaks at an optimal velocity (300 pixel/s). The STMD also exhibits high responses to the objects whose velocities range from 100 to 500 pixel/s. Fig. 13(c) and (d) display the output of the STMD when changing the width and height of the object, which indicate that the STMD prefers moving objects whose widths and heights are smaller than 15 pixels.

These characteristics of the STMD revealed in Fig. 13(a)-(d), are called Weber contrast sensitivity, velocity selectivity, width selectivity and height selectivity, respectively, which have been already found in the STMD neurons in biological research [7], [9], [10].

### C. Effectiveness of the Contrast Pathway

In the proposed visual system model, we design a contrast pathway and incorporate it with the motion pathway to discriminate small targets from fake features. To validate its effectiveness, we first compare the performance of the STMD+ with and without the contrast pathway. Then we conduct a performance comparison between the developed STMD+ and two baseline models including ESTMD [12] and DSTMD [13]. The testing setups are detailed as follows: the input image sequence is presented in Fig. 14(a), which displays a small target moving against the cluttered background; the background is moving from left to right and its velocity is 250 pixel/s; the luminance, size and velocity of small target are equal to 0,  $5 \times 5$  pixels and 250 pixel/s, respectively; the

TABLE II  
DETECTION RATE ( $D_R$ ) AND FALSE ALARM RATE ( $F_A$ ) OF THE STMD+ WITH AND WITHOUT THE CONTRAST PATHWAY UNDER DIFFERENT DETECTION THRESHOLDS  $\beta$ .

Threshold $\beta$	Without*		With#	
	$D_R$	$F_A$	$D_R$	$F_A$
150	<b>0.85</b>	<b>27.70</b>	<b>0.85</b>	<b>0</b>
250	0.74	19.11	0.74	0
350	0.60	12.87	0.60	0
450	0.50	6.88	0.50	0

\* The STMD+ without the contrast pathway.

# The STMD+ with the contrast pathway.

TABLE III  
STANDARD DEVIATIONS OF THE DIRECTION CONTRAST.

	Standard deviation			
	$\phi = 0$	$\phi = \frac{\pi}{4}$	$\phi = \frac{\pi}{2}$	$\phi = \frac{3\pi}{4}$
Small target	<b>36.10</b>	<b>38.17</b>	<b>31.29</b>	<b>42.86</b>
Fake feature	1.38	3.56	3.88	2.54

motion trace of the small target during time period  $[0, 1000]$  ms is illustrated in Fig. 14(b).

Fig. 15(a)-(d) displays the motion traces detected by the STMD+ without the contrast pathway under different detection thresholds  $\beta$ . As can be seen, these detection results all contain numerous fake features. When increasing the detection threshold, the detected fake features will decrease while the detected motion trace becomes more incomplete. After applying the contrast pathway, the fake features are all filtered out even under different detection thresholds (see Fig. 15(e)-(h)). The specific detection rate ( $D_R$ ) and false alarm rate ( $F_A$ ) are presented Table II.

Fig. 16 demonstrates the motion traces detected by the ESTMD, DSTMD and STMD+, where the detection rates ( $D_R$ ) of the three models are all set to 0.85 for fair comparison. As can be seen, the detection results of the ESTMD and DSTMD are seriously contaminated by a number of fake features, whereas that of the STMD+ is noiseless.

To reveal the role of the contrast pathway, we analyze the directional contrast on two motion traces chosen from Fig. 15(a), where one is the small target motion trace, and the other is a randomly selected fake feature trace. Fig. 17 presents the directional contrast on these two motion traces. Note that each motion trace has four directional contrast along four directions  $\phi \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ . As shown in Fig. 17(a), the directional contrast on the motion trace caused by the small target displays significant changes over time. In contrast, the directional contrast of the fake feature trace remains almost unchanged with respect to time (see Fig. 17(b)). The calculated standard deviations of the directional contrast on these two motion traces are listed in Table III, where the sample number  $m$  is equal to 1000. Obviously, the standard deviations of the small target are much larger than those of the fake feature, suggesting that the small target can be discriminated from fake features by comparing their standard deviations.

We further study the relationship of the standard deviations with regard to the sample number  $m$  (see Fig. 18). As it

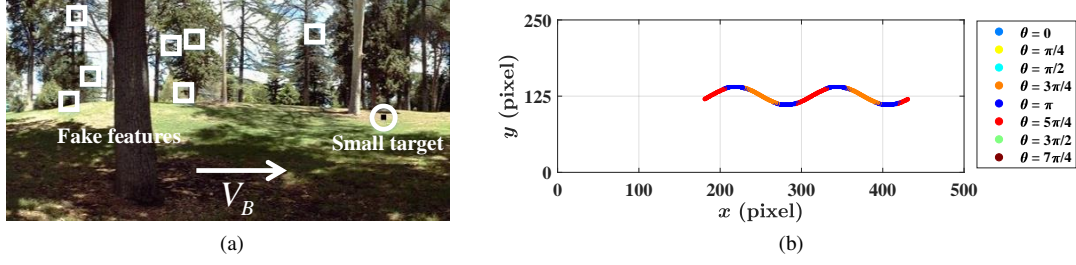


Fig. 14. (a) Representative frame of the input image sequence. A small target (the small black block) highlighted by the circle, is moving against the cluttered background. The background which contains a number of fake features, is also moving from left to right where arrow  $V_B$  denotes the background motion direction. (b) Motion trace of the small target during time period  $[0, 1000]$  ms, i.e., ground truth. In this subplot, color represents motion direction  $\theta$  of the small target.

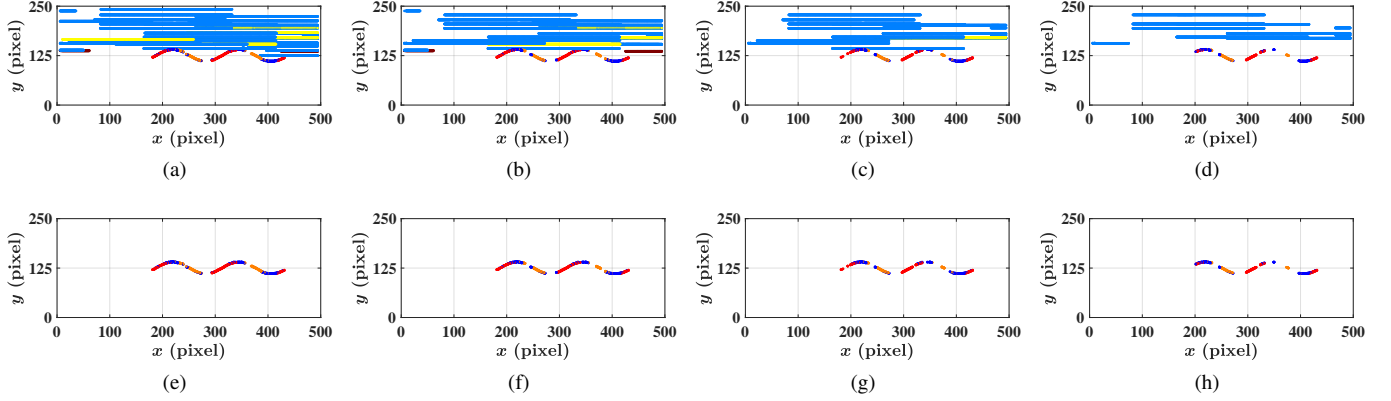


Fig. 15. (a)-(d) Motion traces detected by the STMD+ without the contrast pathway under different detection thresholds  $\beta$  which are set as 150, 250, 350 and 450, respectively. (e)-(h) Motion traces detected by the STMD+ with the contrast pathway under different detection thresholds  $\beta$  which are set as 150, 250, 350 and 450, respectively.

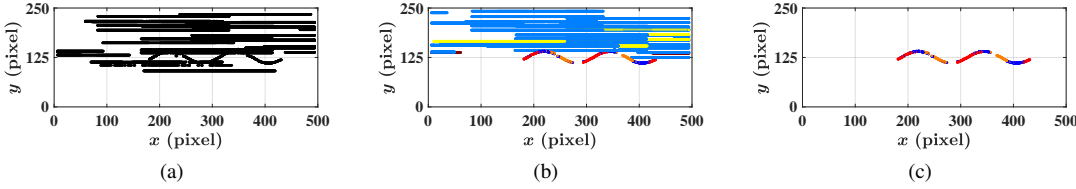


Fig. 16. Motion traces detected by the ESTMD, DSTMD and STMD+. For fair comparison, the three models have fixed detection rates ( $D_R = 0.85$ ). (a) ESTMD. (b) DSTMD. (c) STMD+. Since the ESTMD cannot detect motion direction, its outputs are all shown in black color.

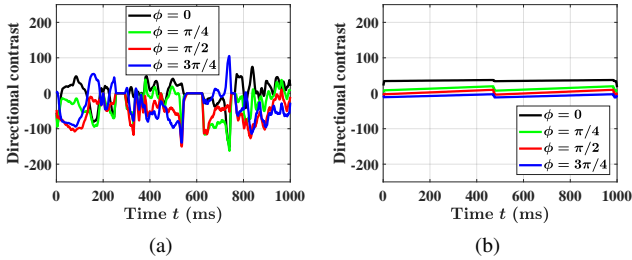


Fig. 17. (a) Directional contrast on the motion trace caused by the **small target**. (b) Directional contrast on the motion trace caused by the **fake feature**. In each subplot, the directional contrast along four directions  $\phi \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$  is presented.

is shown, the standard deviations of the small target exhibit a sharp rise when the sample number increases from 0 to 200. With the continuous growth of the sample number, the standard deviations tend to be stable. Similarly, the standard deviations of the fake feature become stable as the increase of the sample number. Above results indicate that a certain number of samples which is at least greater than 200, is needed to obtain stable standard deviations.

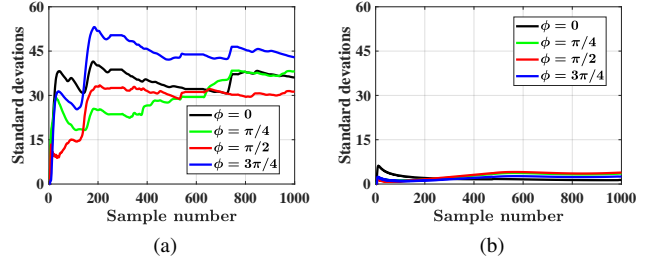


Fig. 18. Standard deviations under different sample numbers. (a) Standard deviations of the **small target**. (b) Standard deviations of the **fake feature**.

#### D. Comparison on Synthetic and Real Datasets

In this section, six groups of synthetic image sequences are first utilized to evaluate the performance of the proposed model in terms of different target velocities, target sizes, target luminance, background velocities, background motion directions and background images. The details of the synthetic image sequences are listed in Table IV. Then the proposed model is further tested on the real dataset (STNS dataset [19]). The performance comparison between the proposed STMD+

TABLE IV  
DETAILS OF THE INITIAL IMAGE SEQUENCE AND SIX GROUPS OF IMAGE SEQUENCES. COMPARING TO THE INITIAL IMAGE SEQUENCE, GROUP 1 TO 6 ARE COMPOSED OF IMAGE SEQUENCES WITH DIFFERENT PARAMETERS.

Parameter	Initial sequence	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Target velocity (pixel/s)	250	0 ~ 500	250	250	250	250	250
Target size (pixel $\times$ pixel)	5 $\times$ 5	5 $\times$ 5	1 $\times$ 1 ~ 12 $\times$ 12	5 $\times$ 5	5 $\times$ 5	5 $\times$ 5	5 $\times$ 5
Target luminance	0	0	0	0 ~ 75	0	0	0
Background velocity (pixel/s)	250	250	250	250	0 ~ 500	0 ~ 500	250
Background motion direction	rightward	rightward	rightward	rightward	rightward	leftward	rightward
Background Image	Fig.14(a)	Fig.14(a)	Fig.14(a)	Fig.14(a)	Fig.14(a)	Fig.14(a)	Fig.20(a) ~ (c)

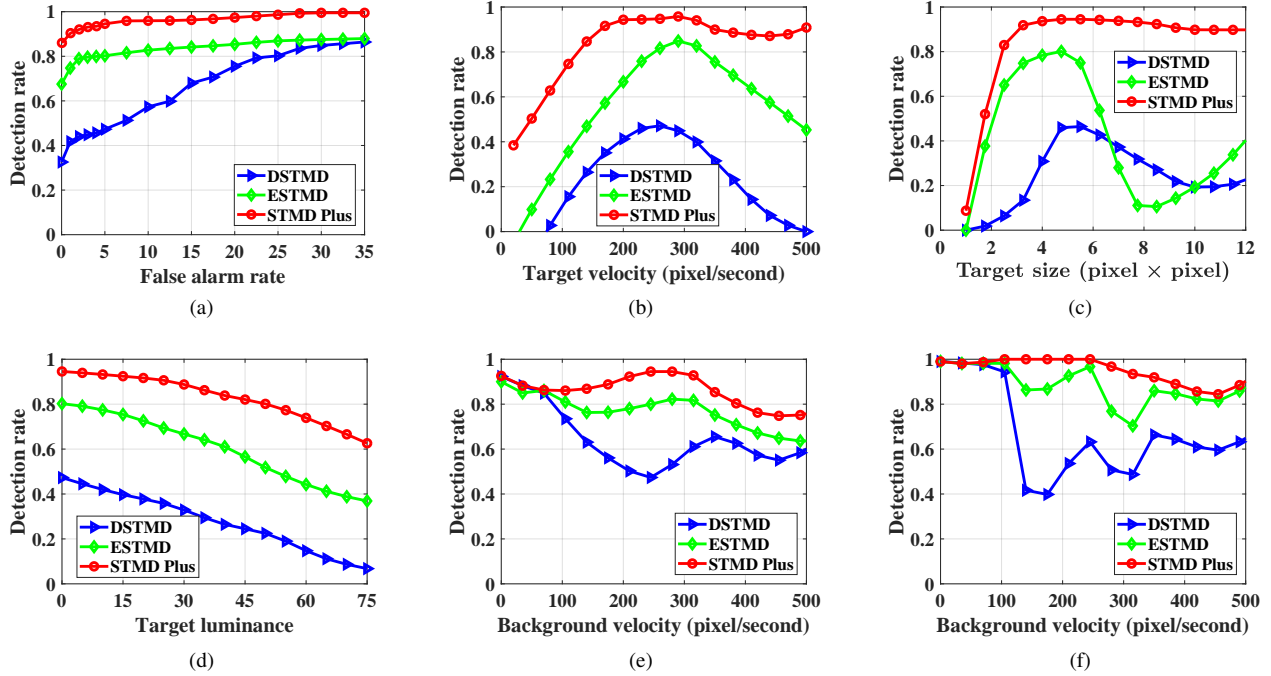


Fig. 19. (a) Receiver operating characteristic (ROC) curves of the three models for the initial image sequence. (b)-(f) Detection rates of the three models for the Group 1-5. For fair comparison, the three models have fixed false alarm rate ( $F_A = 5$ ). (b) Group 1, different target velocities. (c) Group 2, different target sizes. (d) Group 3, different target luminance. (e) Group 4, different background velocities (in rightward motion). (f) Group 5, different background velocities (in leftward motion).

and two baseline models (namely, ESTMD and DSTMD), is also conducted.

Fig. 19(a) shows the receiver operating characteristics (ROC) curves of the three models for the initial synthetic image sequence. It can be seen that the STMD+ has better performance than the baseline models. More precisely, the STMD+ has higher detection rates ( $D_R$ ) compared to the baseline models while the false alarm rates ( $F_A$ ) are low. Fig. 19(b)-(d) display the detection rates of the three models for the Group 1 to 5, where the false alarm rates of the three models are all equal to 5 for fair comparison. From Fig. 19(b) and (c), we can see that the STMD+ significantly outperforms the baseline models. The STMD+ has higher detection rates than the baseline models for different target velocities and sizes. The detection rate of the STMD+ remains stable when the target velocity (or size) ranges from 200 to 500 pixel/s (or from 4  $\times$  4 to 12  $\times$  12 pixel  $\times$  pixel). In contrast, the detection rates of the two baseline models significantly decrease after reach the maximum points. As it shown in Fig. 19(d), the STMD+

consistently performs best under different target luminance. It is worthy to note that the detection rates of the three models all decrease with the increase of target luminance. In Fig. 19(e) and (f), we can see that the STMD+ has the better performance than the baseline models under different background velocities and directions.

Fig. 20 presents the ROC curves of the three models for the Group 6. As can be seen, the STMD+ outperforms the baseline models in different backgrounds. Note that the three models all perform well in Fig. 20(a). Their detection rates are all close to 1 when the false alarm rates are low, and show small differences. This is because the background is much more homogeneous and contains less fake features. However, in more cluttered backgrounds such as Fig. 20(b) and (c), the STMD+ has a much better performance than the other two models.

We further tested the developed model on the publicly available STNS dataset [19]. Fig. 21 illustrates the ROC curves of the three models for the six real image sequences, where

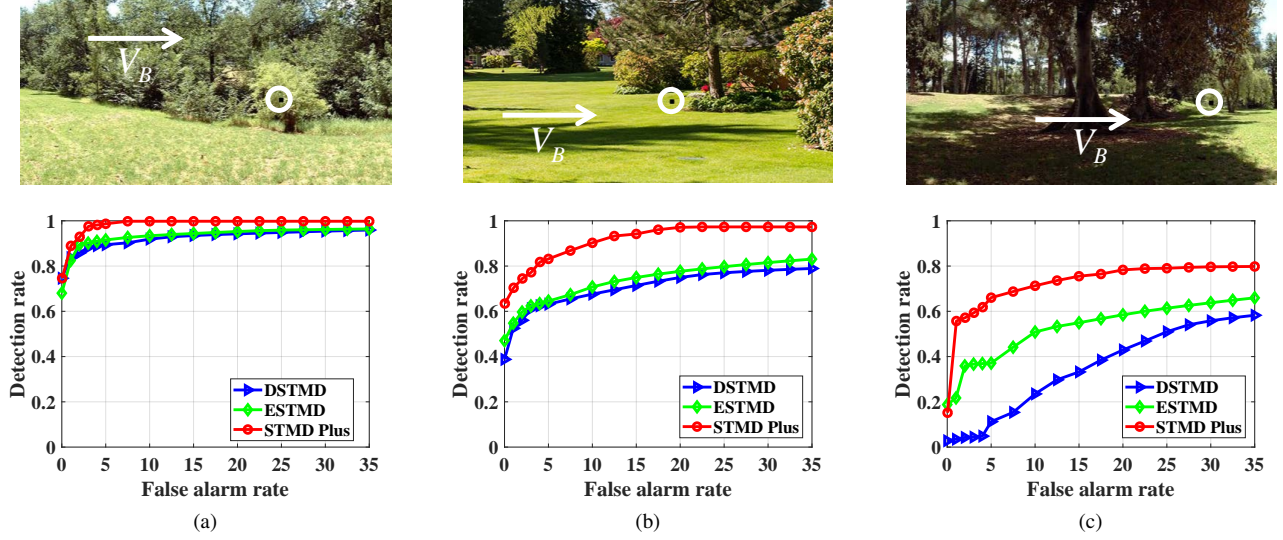


Fig. 20. Background images and receiver operating characteristic (ROC) curves of the three models for the Group 6, different backgrounds.

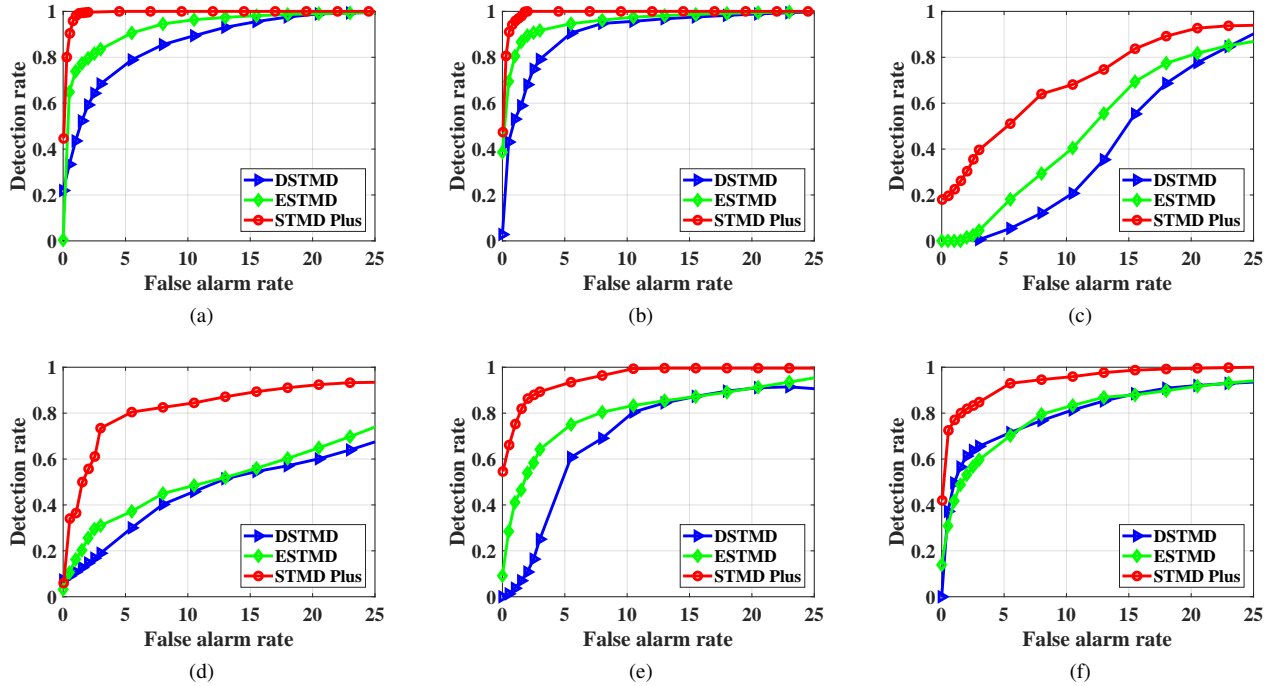


Fig. 21. Receiver operating characteristic (ROC) curves of the three models for the six real image sequences. (a) Real image sequence 1 (STNS-4). (b) Real image sequence 2 (STNS-15). (c) Real image sequence 3 (STNS-16). (d) Real image sequence 4 (STNS-18). (e) Real image sequence 5 (STNS-22). (f) Real image sequence 6 (STNS-25).

the numbers of these six image sequences in the STNS dataset are 4, 15, 16, 18, 22 and 25, respectively. As it is shown in the six subplots, the detection rates of the STMD+ are higher than those of two baseline models when the false alarm rates are given. That is, the STMD+ obtains the best performance for all six real sequences, which means that the STMD+ can work more stably for different cluttered backgrounds and target types.

## V. CONCLUSION

In this paper, we have proposed a visual system model (STMD+) for small target motion detection in cluttered backgrounds. The visual system contains two parallel information pathways and is capable of discriminating small targets from fake features. The first pathway called motion pathway, is intended to locate all small moving objects by calculating luminance changes over time at each pixel. The second pathway called contrast pathway, is designed to capture the directional contrast by computing luminance changes of each pixel along



different directions. The mushroom body is introduced to fuse the two types of information from the two pathways. Finally, small targets are distinguished from fake features by comparing the standard deviations of the directional contrast on their motion traces. Comprehensive evaluation on the synthetic and real datasets, and comparisons with the existing STMD models demonstrate the effectiveness of the proposed visual system in filtering out fake features and improving detection rates. In the future, we will investigate the self-adaptability of the proposed visual system in different environments to further improve the robustness.

## APPENDIX

For demonstration of actual implementations, we attach pseudo-code form of the STMD Plus (see Algorithm 1). We further briefly discuss the complexity of the proposed method for small target motion detection. As shown in Algorithm 1, the computational time of our method mainly consists of four parts: the ommatidia, the motion pathway, the contrast pathway and the mushroom body.

The computational complexity of the ommatidia is essentially determined by a 2-D spatial convolution of the input image with a Gaussian kernel (see Equation (1)), which can be implemented in  $O(k^2mn)$  time for an  $m \times n$  input image and a  $k \times k$  kernel.

In the motion pathway, the LMC output can be regarded as the difference of two Gamma convolutions (see Equation (3)-(5)). Since the temporal Gamma convolution needs  $O(lmn)$  cost where  $l$  is the length of the Gamma kernel, the computational complexity of the LMC scales with  $O(2lmn)$ . Similarly, the total cost of the four medulla neurons is about  $O(2lmn + 2mn)$ . According to (11), the computational complexity of the STMD is  $O(2mn)$  for each preferred direction, so its entire cost grows like  $O(2dmn)$  where  $d$  denotes the number of the preferred directions. Finally, the lateral inhibition mechanism which is implemented by a 2-D convolution (see Equation (12)), needs  $O(k^2mn)$  cost. Thus the entire computational complexity of the motion pathway is  $O((k^2 + 4l + 2d + 2)mn)$ .

In the contrast pathway, the directional contrast of each pixel along different spatial directions is calculated by convolving the ommatidium output with directional derivative operators (see Equation (17)). Since the 2-D spatial convolution needs  $O(k^2mn)$  cost for each spatial direction, the entire computational complexity of the contrast pathway is  $O(k^2dmn)$ .

In the mushroom body, the nearest neighbor of each detected object is calculated for recording motion trace, which can be implemented in  $O(p \log p)$  time [53] where  $p$  is the number of the detected objects. In addition, the cost of standard deviation calculation is around  $O(rp)$  where  $r$  represents the sample number. So the entire computational complexity of the mushroom body is around  $O(p \log p + rp)$ .

Based on the aforementioned analysis, the entire computational complexity of the proposed STMD Plus is around  $O(N(2k^2 + k^2d + 4l + 2d + 2)mn + N(\log p + r)p)$  where  $N$  stands for the number of input images.

## Algorithm 1 Detection Process of the STMD Plus

---

**Input:** Image sequence  $\{I_1, I_2, \dots, I_N\}$ , where  $I_i \in \mathbb{R}^{m \times n}$ .  
**Output:** Positions of small moving targets in each input image.

---

```

1: for each input image do
2:   // Ommatidia
3:   Calculate the output of the ommatidium via (1);
4:   // Motion Pathway
5:   Calculate the output of the LMC via (5);
6:   Calculate the outputs of the medulla neurons via (6)-(9);
7:   Calculate the output of the STMD via (11);
8:   Calculate the laterally inhibited output via (12);
9:   // Contrast Pathway
10:  Calculate the output of the AMC via (16);
11:  Calculate the output of the T1 neuron via (17);
12:  // Mushroom Body
13:  Calculate motion traces of the detected objects via (20);
14:  for each motion trace do
15:    Calculate the directional contrast of the motion trace via (21);
16:    Calculate the standard deviations ( $SD$ ) of the directional contrast on the motion trace;
17:    if  $SD > \text{threshold}$  then
18:      the detected object is a small target;
19:    else
20:      the detected object is a fake feature.
21:    end if
22:  end for
23: end for

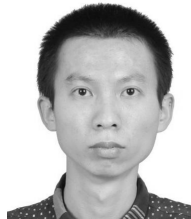
```

---

## REFERENCES

- [1] S. Yue, F. C. Rind, M. S. Keil, J. Cuadri, and R. Stafford, "A bio-inspired visual collision detection mechanism for cars: Optimisation of a model of a locust neuron to a novel environment," *Neurocomputing*, vol. 69, no. 13, pp. 1591–1598, Feb. 2006.
- [2] S. Yue and F. C. Rind, "Collision detection in complex dynamic scenes using an lgmd-based visual neural network with feature enhancement," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 705–716, May 2006.
- [3] M. Woźniak and D. Połap, "Adaptive neuro-heuristic hybrid model for fruit peel defects detection," *Neural Netw.*, vol. 98, pp. 16–33, Feb. 2018.
- [4] C. Yan, H. Xie, J. Chen, Z. Zha, X. Hao, Y. Zhang, and Q. Dai, "A fast uyghur text detector for complex background images," *IEEE Trans. Multimed.*, vol. 20, no. 12, pp. 3389–3398, Dec. 2018.
- [5] C. Yan, H. Xie, S. Liu, J. Yin, Y. Zhang, and Q. Dai, "Effective uyghur language text detection in complex background images for traffic prompt identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 220–229, Jan. 2018.
- [6] C. Yan, H. Xie, D. Yang, J. Yin, Y. Zhang, and Q. Dai, "Supervised hash coding with deep neural network for environment perception of intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 284–295, Jan. 2018.
- [7] P. D. Barnett, K. Nordström, and D. C. O'Carroll, "Retinotopic organization of small-field-target-detecting neurons in the insect visual system," *Curr. Biol.*, vol. 17, no. 7, pp. 569–578, Apr. 2007.
- [8] M. Mischiati, H.-T. Lin, P. Herold, E. Imler, R. Olberg, and A. Leonardo, "Internal models direct dragonfly interception steering," *Nature*, vol. 517, no. 7534, pp. 333–338, Jan. 2015.
- [9] M. F. Keleş and M. A. Frye, "Object-detecting neurons in drosophila," *Curr. Biol.*, vol. 27, no. 5, pp. 680–687, Mar. 2017.
- [10] K. Nordström, P. D. Barnett, and D. C. O'Carroll, "Insect detection of small targets moving in visual clutter," *PLoS Biol.*, vol. 4, no. 3, p. e54, Feb. 2006.
- [11] K. Nordström, "Neural specializations for small target detection in insects," *Curr. Opin. Neurobiol.*, vol. 22, no. 2, pp. 272–278, Apr. 2012.

- [12] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, "A model for the detection of moving targets in visual clutter inspired by insect physiology," *PLoS One*, vol. 3, no. 7, pp. 1–11, Jul. 2008.
- [13] H. Wang, J. Peng, and S. Yue, "A directionally selective small target motion detecting visual neural network in cluttered backgrounds," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2018.2869384.
- [14] L. Freifeld, D. A. Clark, M. J. Schnitzer, M. A. Horowitz, and T. R. Clandinin, "Gabaergic lateral interactions tune the early stages of visual processing in drosophila," *Neuron*, vol. 78, no. 6, pp. 1075–1089, Jun. 2013.
- [15] R. Behnia, D. A. Clark, A. G. Carter, T. R. Clandinin, and C. Desplan, "Processing properties of on and off pathways for drosophila motion detection," *Nature*, vol. 512, no. 7515, p. 427, Aug. 2014.
- [16] Q. Fu, C. Hu, J. Peng, and S. Yue, "Shaping the collision selectivity in a looming sensitive neuron model with parallel on and off pathways and spike frequency adaptation," *Neural Netw.*, vol. 106, pp. 127–143, Oct. 2018.
- [17] B. Hu, S. Yue, and Z. Zhang, "A rotational motion perception neural network based on asymmetric spatiotemporal visual information processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2803–2821, Nov 2016.
- [18] S. Yue and F. C. Rind, "Redundant neural vision systems-competing for collision recognition roles," *IEEE Trans. Auton. Mental Develop.*, vol. 5, no. 2, pp. 173–186, Apr. 2013.
- [19] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Performance of an insect-inspired target tracker in natural conditions," *Bioinspir. & Biomim.*, vol. 12, no. 2, p. 025006, Feb. 2017.
- [20] A. L. Steckl, W. A. Ribi, and E. J. Warrant, "Adaptations for nocturnal and diurnal vision in the hawkmoth lamina," *J. Comp. Neurol.*, vol. 524, no. 1, p. 160, Jul. 2016.
- [21] Z. Riveraalvidrez, I. Lin, and C. M. Higgins, "A neuronally based model of contrast gain adaptation in fly motion vision," *Visual Neurosci.*, vol. 28, no. 5, p. 419, Aug. 2011.
- [22] N. Lessios, R. L. Rutowski, J. H. Cohen, M. E. Sayre, and N. J. Strausfeld, "Multiple spectral channels in branchiopods. i. vision in dim light and neural correlates," *J. Exp. Biol.*, pp. jeb-165 860, Apr. 2018.
- [23] Y.-J. Lee, H. O. Jönsson, and K. Nordström, "Spatio-temporal dynamics of impulse responses to figure motion in optic flow neurons," *PLoS One*, vol. 10, no. 5, pp. 1–16, May 2015.
- [24] J. Li, J. Lindemann, and M. Egelhaaf, "Local motion adaptation enhances the representation of spatial structure at emd arrays," *PLOS Comput. Biol.*, vol. 13, no. 12, pp. 1–23, Dec. 2017.
- [25] S. D. Wiederman and D. C. O'Carroll, "Biologically inspired feature detection using cascaded correlations of off and on channels," *J. Artif. Intell. Soft Comput. Res.*, vol. 3, no. 1, pp. 5–14, Dec. 2013.
- [26] B. Hassenstein and W. Reichardt, "Systemtheoretische analyse der zeit-, reihenfolgen-und vorzeichenbewertung bei der bewegungsperzeption des rüsselkäfers chlorophanus," *Zeitschrift für Naturforschung B*, vol. 11, no. 9-10, pp. 513–524, Oct. 1956.
- [27] H. Eichner, M. Joesch, B. Schnell, D. F. Reiff, and A. Borst, "Internal structure of the fly elementary motion detector," *Neuron*, vol. 70, no. 6, pp. 1155–1164, Jun. 2011.
- [28] H. Wang, J. Peng, and S. Yue, "An improved lptc neural model for background motion direction estimation," in *Proc. IEEE Conf. ICDL-EpiRob*, 2017, pp. 47–52.
- [29] D. A. Clark, L. Bursztyn, M. A. Horowitz, M. J. Schnitzer, and T. R. Clandinin, "Defining the computational structure of the motion detector in drosophila," *Neuron*, vol. 70, no. 6, pp. 1165–1177, Jun. 2011.
- [30] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: a survey," *Comput. Vis. Image Underst.*, vol. 134, pp. 1–21, May 2015.
- [31] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1726–1740, Jul. 2017.
- [32] Z. Li, G. Zhao, S. Li, H. Sun, R. Tao, X. Huang, and Y. J. Guo, "Rotation feature extraction for moving targets based on temporal differencing and image edge detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1512–1516, Oct. 2016.
- [33] Y. Ren, C.-S. Chua, and Y.-K. Ho, "Motion detection with nonstationary background," *Mach. Vis. Appl.*, vol. 13, no. 5-6, pp. 332–343, Mar. 2003.
- [34] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4996–5009, Sep. 2013.
- [35] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognit.*, vol. 58, pp. 216–226, Oct. 2016.
- [36] X. Bai and Y. Bi, "Derivative entropy-based contrast measure for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2452–2466, Apr. 2018.
- [37] C. Hu, F. Arvin, C. Xiong, and S. Yue, "Bio-inspired embedded vision system for autonomous micro-robots: the lgmd case," *IEEE Trans. Cogn. Develop. Syst.*, vol. 9, no. 3, pp. 241–254, Sep. 2016.
- [38] G. Indiveri and R. Douglas, "Neuromorphic vision sensors," *Science*, vol. 288, no. 5469, pp. 1189–1190, May 2000.
- [39] F. C. Rind, R. D. Santer, J. M. Blanchard, and P. F. Verschure, "Locusts looming detectors for robot sensors," in *Sensors and sensing in biology and engineering*. Springer, 2003, pp. 237–250.
- [40] W. He and Y. Dong, "Adaptive fuzzy neural network control for a constrained robot using impedance learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1174–1186, Apr. 2018.
- [41] D. Polap, M. Woźniak, W. Wei, and R. Damaševičius, "Multi-threaded learning control mechanism for neural networks," *Future Gener. Comput. Syst.*, vol. 87, pp. 16–34, Oct. 2018.
- [42] P. Ardin, F. Peng, M. Mangan, K. Lagogiannis, and B. Webb, "Using an insect mushroom body circuit to encode route memory in complex natural environments," *PLoS Comput. Biol.*, vol. 12, no. 2, pp. 1–12, Feb. 2016.
- [43] B. Webb and A. Wystrach, "Neural mechanisms of insect navigation," *Curr. Opin. Insect Sci.*, vol. 15, pp. 27–39, Jun. 2016.
- [44] E. J. Warrant, "Matched filtering and the ecology of vision in insects," in *The Ecology of Animal Senses*. Springer, Dec. 2016, pp. 143–167.
- [45] S.-y. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston *et al.*, "A visual motion detection circuit suggested by drosophila connectomics," *Nature*, vol. 500, no. 7461, p. 175, Aug. 2013.
- [46] R. Behnia and C. Desplan, "Visual circuits in flies: beginning to see the whole picture," *Curr. Opin. Neurobiol.*, vol. 34, pp. 125–132, Oct. 2015.
- [47] B. De Vries and J. C. Principe, "A theory for neural networks with time delays," in *Proc. NIPS*, 1990, pp. 162–168.
- [48] S. Yamaguchi and M. Heisenberg, "Photoreceptors and neural circuitry underlying phototaxis in insects," *Fly*, vol. 5, no. 4, pp. 333–336, Dec. 2011.
- [49] S. M. Rogers and S. R. Ott, "Differential activation of serotonergic neurons during short-and long-term gregarization of desert locusts," *Proc. Royal Soc. B*, vol. 282, no. 1800, p. 20142062, Feb. 2015.
- [50] W. T. Freeman, E. H. Adelson *et al.*, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
- [51] W.-C. Zhang and P.-L. Shui, "Contour-based corner detection via angle difference of principal directions of anisotropic gaussian directional derivatives," *Pattern Recognit.*, vol. 48, no. 9, pp. 2785–2797, Sep. 2015.
- [52] A. D. Straw, "Vision egg: an open-source library for realtime visual stimulus generation," *Front. Neuroinf.*, vol. 2, no. 4, Nov. 2008.
- [53] A. Mariello and R. Battiti, "Feature selection based on the neighborhood entropy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6313–6322, Dec. 2018.



**Hongxin Wang** received the B.Sc. and M.Sc. degrees in mathematics and applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the School of Computer Science, University of Lincoln, Lincoln, U.K.

His current research interests include small target detection models, artificial intelligence, and computational neuroscience.



**Jigen Peng** received the B.Sc. degree from Jiangxi University, Nanchang, China, in 1989, and the M.Sc. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1992 and 1998, respectively.

He was with Xi'an Jiaotong University as a Lecturer from 1992 to 1998, an Associate Professor from 1998 to 2002, and a Professor from 2002 to 2018. He is currently a Professor with the School of Mathematics and Information Science, Guangzhou University, Guangzhou, China. His current research interests include nonlinear functional analysis and

applications, machine learning theory, and sparse information processing.



**Xuqiang Zheng** received the B.Sc. and M.Sc. degrees from the School of Physics and Electronics, Central South University, Hunan, China, in 2006 and 2009, respectively, and Ph.D. degree from the University of Lincoln, Lincoln, U.K., in 2018.

He is currently an Associate Professor with the Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China. From 2010 to 2015, he was a Mixed Signal Engineer with the Institute of Microelectronics, Tsinghua University, Beijing, China. His current research interests include high-

performance A/D converters and highspeed wireline communication systems.



**Shigang Yue** (M'05–SM'17) received the B.Eng. degree from Qingdao University of Technology, Qingdao, China, in 1988, and the M.Sc. and Ph.D. degrees from the Beijing University of Technology (BJUT), Beijing, China, in 1993 and 1996, respectively.

He was with BJUT as a Lecturer from 1996 to 1998 and an Associate Professor from 1998 to 1999. He was an Alexander von Humboldt Research Fellow at the University of Kaiserslautern, Kaiserslautern, Germany, from 2000 to 2001. He

is currently a Professor of computer science with the School of Computer Science, University of Lincoln, Lincoln, U.K. Before joining the University of Lincoln as a Senior Lecturer in 2007 and promoted to Reader in 2010 and Professor in 2012, he held research positions with the University of Cambridge, Cambridge, UK, Newcastle University, Newcastle upon Tyne, UK, and University College London, London, UK, respectively. His current research interests include artificial intelligence, computer vision, robotics, brains and neuroscience, biological visual neural systems, evolution of neuronal subsystems, and their applications, e.g., in collision detection for vehicles, interactive systems, and robotics.

Dr. Yue is a member of the International Neural Network Society, International Society of Artificial Life, and International Symposium on Biomedical Engineering. He is the founding director of the Computational Intelligence Laboratory (CIL), Lincoln, and deputy director of Lincoln Centre for Autonomous Systems (L-CAS), Lincoln. He is also a co-founding director of the Machine Life and Intelligence Research Center, Guangzhou, China. He is the coordinator for several EU FP7/Horizon2020 projects.